Objectives

This algorithm is intended to be used by football enthusiasts and gamblers to help predict the outcome of an upcoming football match based on the results of the previous games. The model takes raw data in form of CSV files representing historical data of football matches and uses Poisson distribution to predict the home team winning probability, probability of a draw or the probability of the away team winning.

Datasets

The datasets used in this model are obtained from https://www.football-data.co.uk/englandm.php and represents historical data of past football games. The data is downloaded from the website or can be directly accessed from the URL using the Python Pandas module.

The datasets have numerous columns but we intend to use HomeTeam, AwayTeam, FTHG and FTAG. The first column is used to represent the home team, the second one is the away team, FTHG is short for full time home goals and FTAG is full time away goals.

Model

We are going to use Poisson distribution to predict the probability of 3 outcomes in this model; the percentage probability of the home team winning, the percentage probability of a draw and the percentage probability of the away team winning. To achieve this, we are going to treat the number of goals the home team scores and the number of goals scored by the away team as independent Poisson events.

The team that scores the most number of goals is the winner and if the number of goals scored by each team is zero then the game ends as a draw, also if the teams score the same number of goals.

Tools

We are going to use python for this project. The following python modules will be used; . Pandas is used to read, manipulate and visualize data, matplotlib is used to visualize data, numpy is used to manipulate mathematical functions on a dataset, seaborn is also used for data visualization and scipy is for scientific and technical computing.